

Data QUEST

Data Quality Testing – DQe Tools

2/28/17

Kari A. Stephens, PhD

Assistant Professor, Psychiatry & Behavioral Sciences

Adjunct Assistant Professor, Biomedical Informatics & Medical Education

WWAMI region Practice & Research Network



- ~58 Primary care WWAMI clinics
 - ~20 data connected clinics
 - CHCs and RHCs
 - Underserved populations
 - Many serving rural populations
 - Collaboration with national network of practice based research networks
 - Data QUEST represents over 250,000 patients
- <https://dataquest.iths.org/>
<https://github.com/WWAMI-DataQuest>

login

Data QUEST

Home

Explore Data

Success Stories

About Us

ITHS

Institute of Translational Health Sciences
Accelerating Research. Improving Health.

Data QUEST, supported by the Institute of Translational Health Sciences, is an electronic health data-sharing architecture across community-based primary care practices in Washington and Idaho.

Data QUEST is designed to provide access to research datasets generated from electronic medical record systems within our primary care community-based practice partner settings to catalyze both regional and national health discoveries.

Clinical Data Repository

Partner 1

Partner 2

Partner 3

Local Sites
Data repository
resides at clinic

Combined Clinical
Data Repository
De-identified for research

Comparative
Effectiveness

Randomized
Control Trials

Cohort
Discovery

Data Sharing
Practice approves
for each project

From cohort discovery to clinical trials to comparative effectiveness research, you can use our innovative data-sharing tools to streamline and enhance your next community-based research study.

Our expert team will work with you to determine how to conduct your project with our community-based partners. We also offer technical assistance to help you define datasets to drive your research.

Please click on the Browse Data button to begin browsing the data types and diagnosis categories contained in the Data QUEST data repository to get to know what data are available and if they suit your research needs.

Browse Data

Contact Us Today

Funding Opportunity

Do you have a research question you think could be answered by analyzing primary care electronic medical record data? If so, we can help by underwriting the cost of data extraction as part of our launch.

To be considered for this opportunity, please email a paragraph describing your research question, intended use of the data, and description of your professional role to Gina Keppel (gakeppel@uw.edu).

Search

QUICK LINKS

Browse Data

Request Data

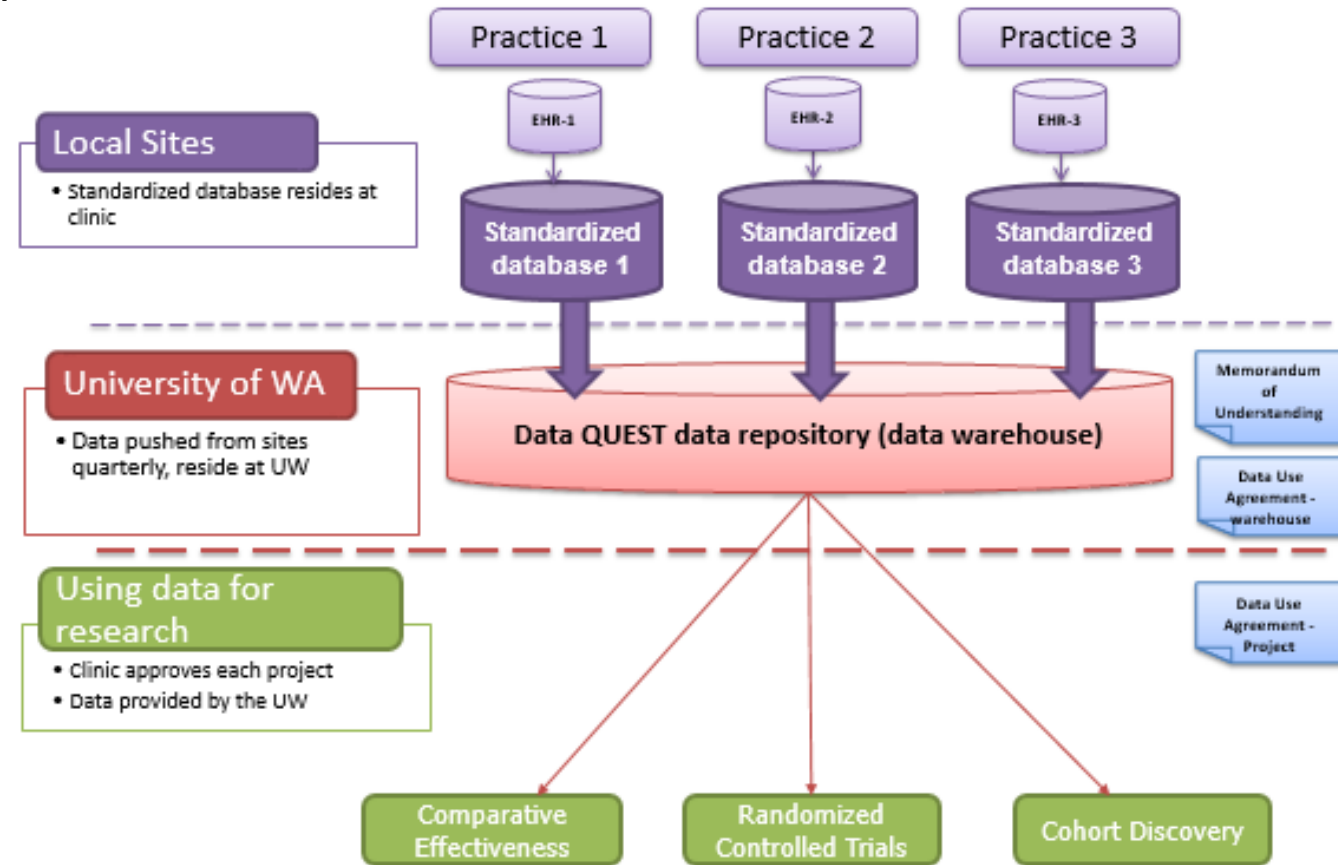
Request a Consult

About ITHS | [ITHS Researcher Resources](#) | [Privacy Statement](#)

Data QUEST

- 20 data-connected clinics in the WPRN
- Represents over 250,000 patients

An electronic health data-sharing architecture across community-based primary care practices in the WPRN



Current Clinical Research Trials

- Team-based Safe Opioid Prescribing – dissemination trial across 6 regional primary care practices (AHRQ)
- Integrating Behavioral Health and Primary Care – large national pragmatic trial across 40 national primary care practices (PCORI)

Network Participation

- PCORNet's Patient-Centered Scalable National Network for Effectiveness Research (pSCANNER) (PCORI)
- Clinical Trials Network: Pacific Northwest Node (NIH/NIDA)
- Accelerating Change and Transformation in Organization and Networks III (ACTION III) partnership, The Quality Commons (AHRQ)
- WWAMI Practice Transformation Network (CMS)
- Diabetes Prevention Registry (CDC)
- Northwest Pharmacogenomic Research Network (NIH/NIGMS)
- DARTNet Practice Benchmarking Registry (industry)
- MOSAIC: Meaningful Outcomes and Science to Advance Innovations Center of Excellence (AHRQ)

Data QUEST: Improving Health in Rural Populations

funded by NIH, AHRQ,
CDC, PCORI, AHRQ,
CMMS, and industry

Common Tables – OMOP V.4

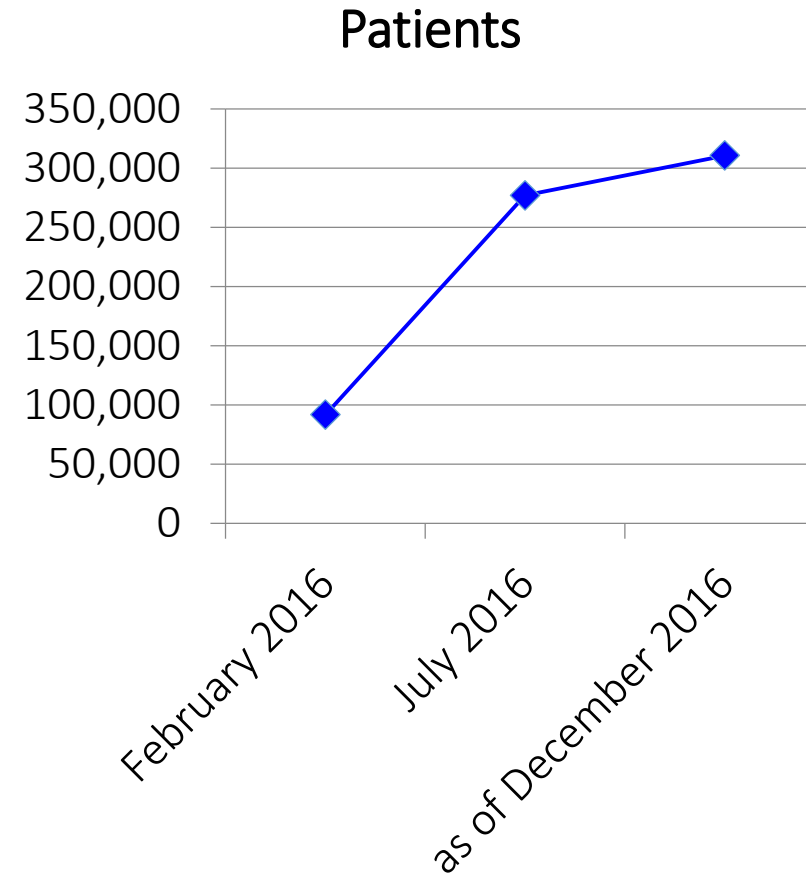
- Care Site
 - Sites at each organization
- Condition Occurrence
 - Encounter associated diagnoses
 - Problem list diagnoses
- Drug Exposure
 - Medications
- Location
 - Patient and site addresses
- Observation
 - Vitals and Labs
 - Past medical history
 - Family history
- Person
 - Patient demographics
- Procedure Occurrence
 - Encounter associated procedures
 - CPT codes
- Visit Occurrence
 - Appointments
 - Encounters

Current UW-hosted Data QUEST Warehouse Patients

310,604 patients in the person table

- 102,330 (33%) at Organization B
- 45,685 (15%) at Organization C
- 27,577 (9%) at Organization N
- 36,001 (12%) at Organization P
- 99,011 (32%) at Organization Y

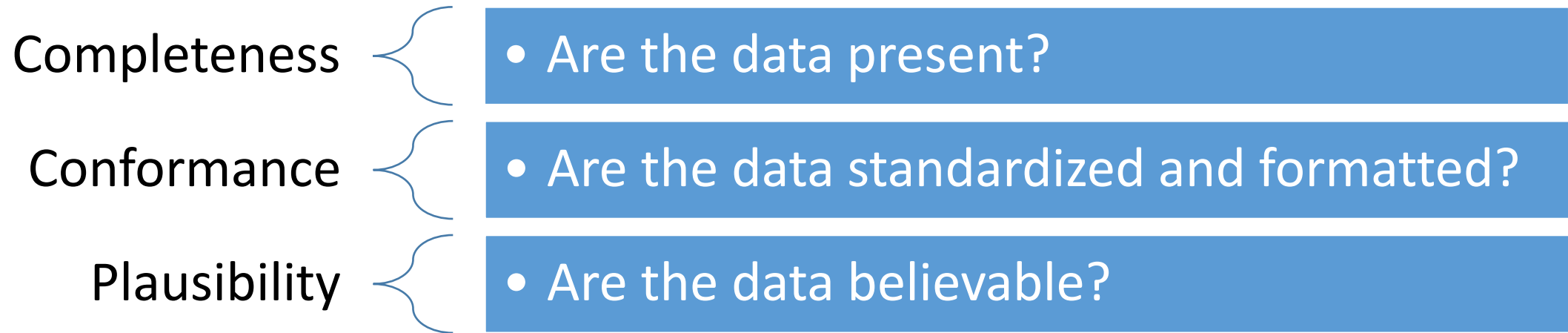
10M encounters



Measuring Data Quality

A new framework...

Operationalizing the framework into: 5 conceptual tests and 17 discrete tests across:



Kahn et al. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. eGEMS, 4, 1244.

<https://www.ncbi.nlm.nih.gov/pubmed/27713905>

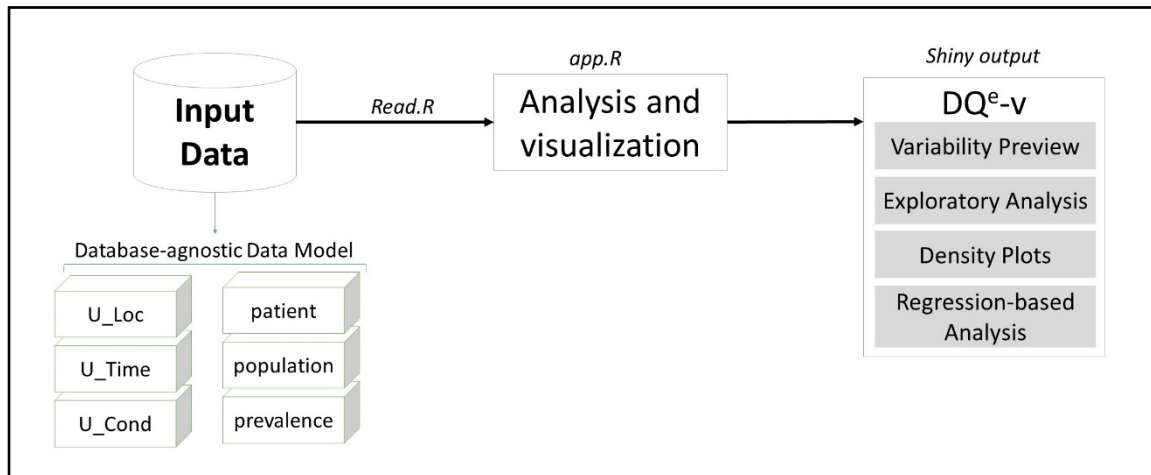
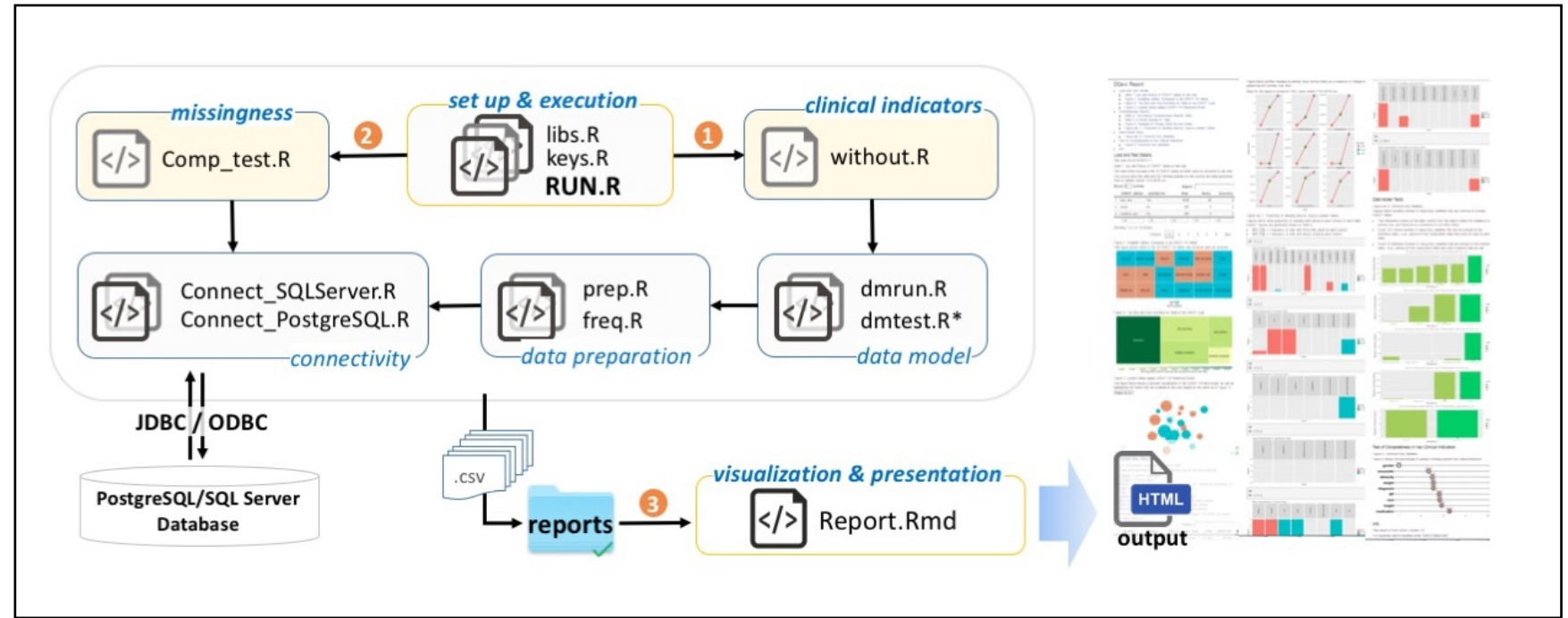
Data Quality Tests

TEST ID	DOMAIN	TEST
C1	COMPLETENESS	Number of Tables Received, Number of Observations, Flag Indicator for the table having actual data
C2	COMPLETENESS	GENDER completeness (denominator and proportion with valid data)
C3	COMPLETENESS	AGE/DOB completeness (denominator and proportion with valid data)
C4	COMPLETENESS	VITALS completeness (denominator and proportion with valid data): Height, Weight, SBP, DBP
C5	COMPLETENESS	LABS completeness (denominator and proportion with valid data): A1c, HDL, LDL, Triglycerides, Total cholesterol
F1	FIDELITY	Check that primary and foreign keys relate properly; High Priority: Person_ID, Visit_Occurrence_ID
F2	FIDELITY	Duplicate patient check in the patient table (Find the same patient with a different patient ID using full name, dob, and gender)
F3	FIDELITY	Visualize codes/values entered for DEMOGRAPHICS (Gender, Race, Ethnicity)
F4	FIDELITY	Visualize YEAR OF BIRTH to help identify errors or missing cohorts
P1	PLAUSIBILITY	Comparison of new load to old load (Number of observations, Number of unique patients, Number of tables with rows)
P2	PLAUSIBILITY	Review of minimum and maximum dates for tables with key dates; High Priority: Visit_Occurrence table
P3	PLAUSIBILITY	How many patients have a year of birth after their visit dates?
P4	PLAUSIBILITY	Check that certain observation types fall into specific ranges
P5	PLAUSIBILITY	Visualize number of visits in a year or across years
P6	PLAUSIBILITY	Visualize type of visit in a year or across years
P7	PLAUSIBILITY	Volume Check: Proportion of patients with visit data and select observation types
P8	PLAUSIBILITY	Logical Constraints Check

DQe Tool Architecture

DQe-c

modular tool
developed in R
statistical language
for assessing
completeness in EHR
data repositories



DQe-v

interactive interface
powered by the shiny
package version
0.13.0 in R

Operationalizing use of DQe tools for data quality testing

- * Data QUEST
- * DARTNet Institute



DQe-c/DQe-v Reports Standard Operating Procedure (SOP)

Version 2 December 2016

TABLE OF CONTENTS

DQe-c Overview.....	2
How to Use the Tables in the DQe-c Report	2
Load and Test Detail Section.....	4
Table 1.List and Status of OMOP Tables in this Load.....	4
Figure 1.Available Tables, Compared to all OMOP V4 Tables.....	4
Figure 2.File Size and Row Numbers by Table in the OMOP Load.....	5
Figure 3.Loaded Tables against OMOP V4 Relational Model.....	6
Completeness Results Section.....	8
Table 2.The Master Completeness Results Table.....	8
Table 3.A Small Sample of Flags.....	9
Figure 4.Changes in Primary Keys across Loads.....	10
Figure Set 1.Proportion of Missing Data by Type in Loaded Tables.....	10
Data Model Tests Section.....	11
Figure Set 2.Common Key Variables.....	11
Test of Completeness in Key Clinical Indicators Section.....	13
Figure 5.Common Key Variables.....	13
DQe-c Aggregated Report Overview.....	14
Figure 1. Overall missingness across OMOP tables and columns.....	14
Figure 2. Overall missingness in key indicators.....	15
DQe-v Overview.....	16
Variability Preview Tab.....	17
Exploratory Analysis Tab.....	18
Density Plot Tab.....	19
Regression-Based Analysis Tab.....	19
Appendix A. Data Quality Tests.....	20

DQe-c and DQe-v Report Flows

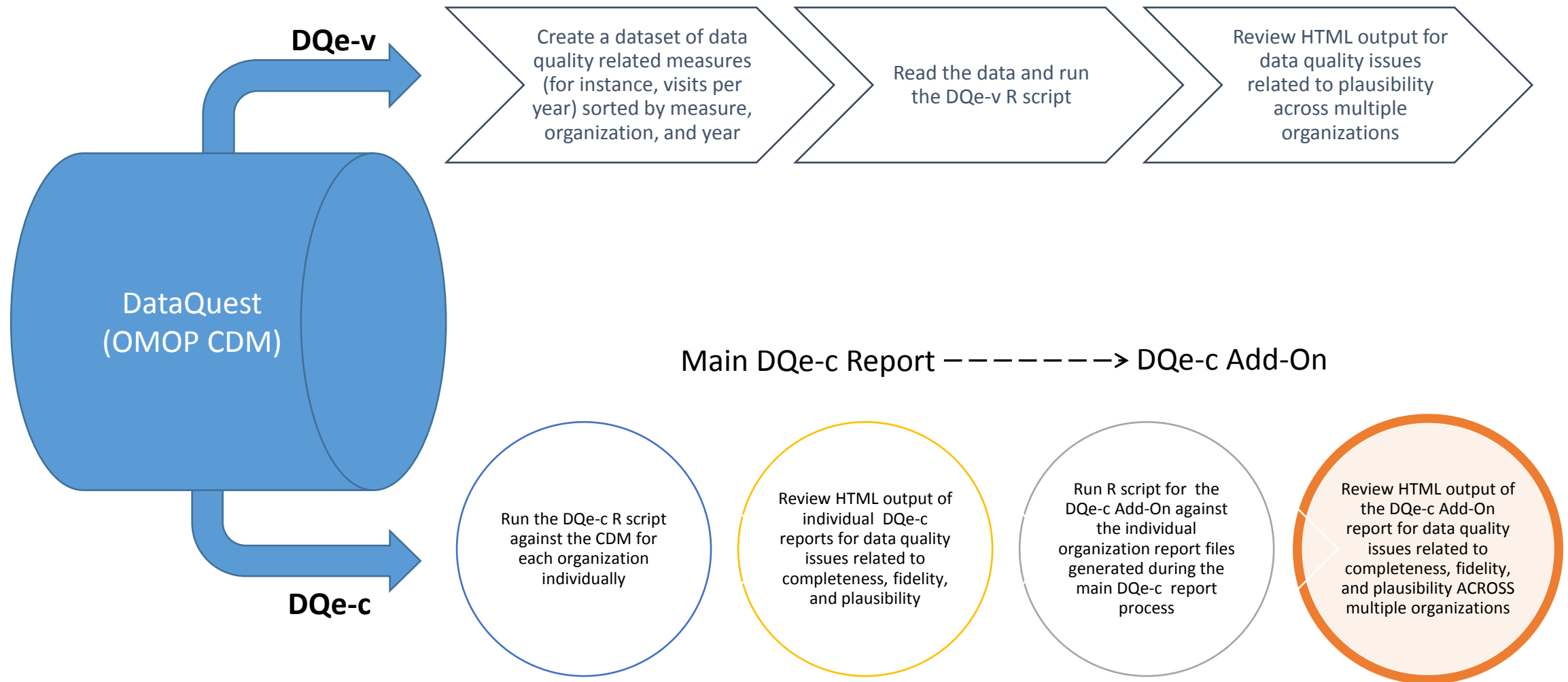



Figure 3. Loaded tables against OMOP V4 Relational Model.

The figure below shows a network visualization of the OMOP V4 data model, as well as highlighting the tables that are available in this load (legend is the same as in Figure 1).

Select by id 

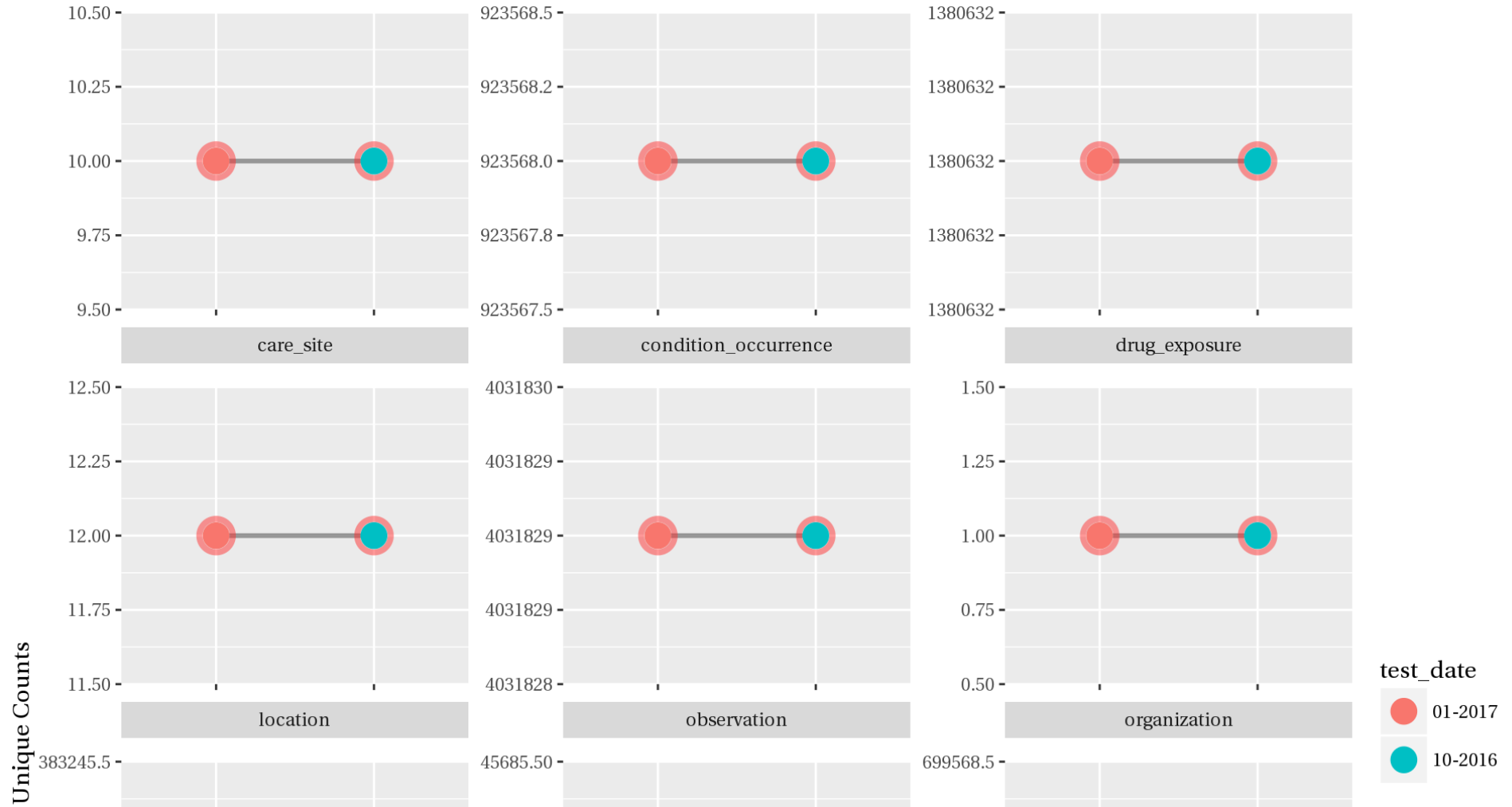
The network's table schemas and key relationships

- Color coated to display “missingness”

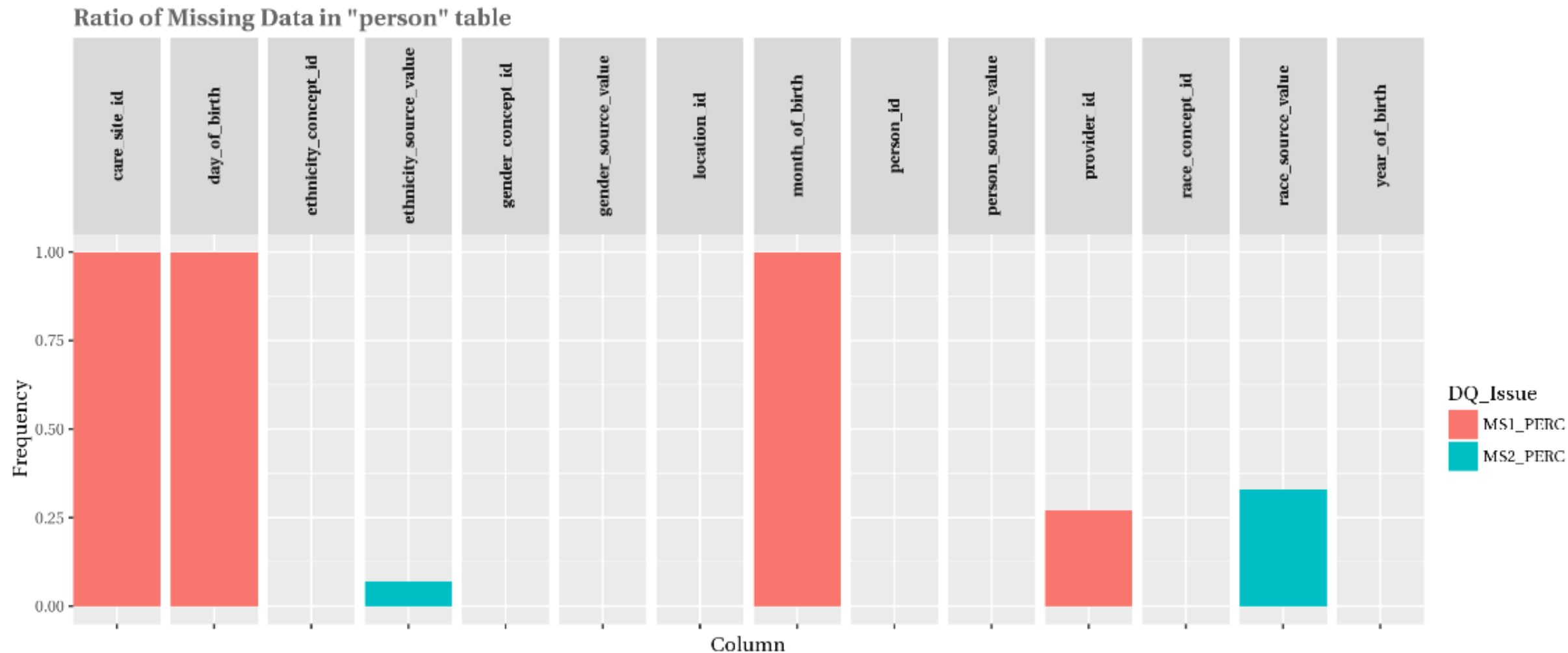


Completeness example:

Number of primary keys for available tables over time

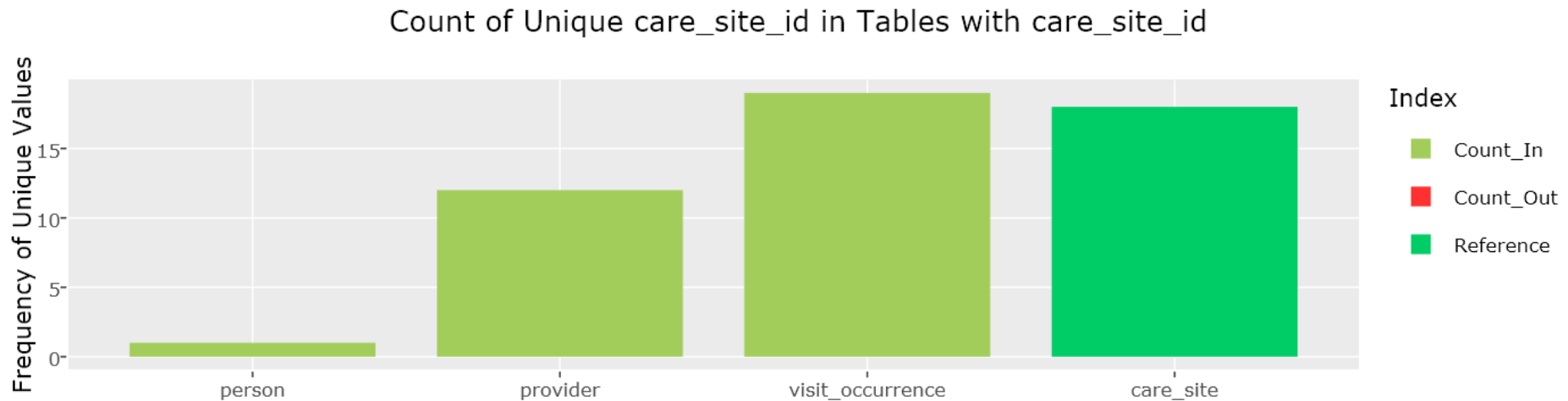
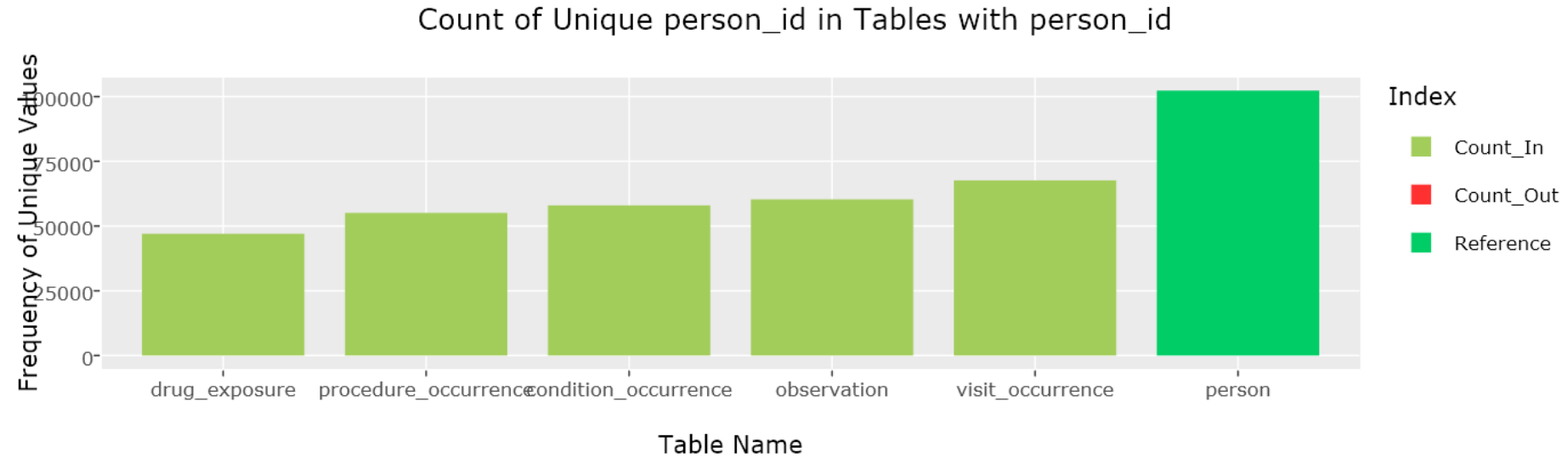


Completeness example: Detailing columns with proportion of missingness (null vs. blank)

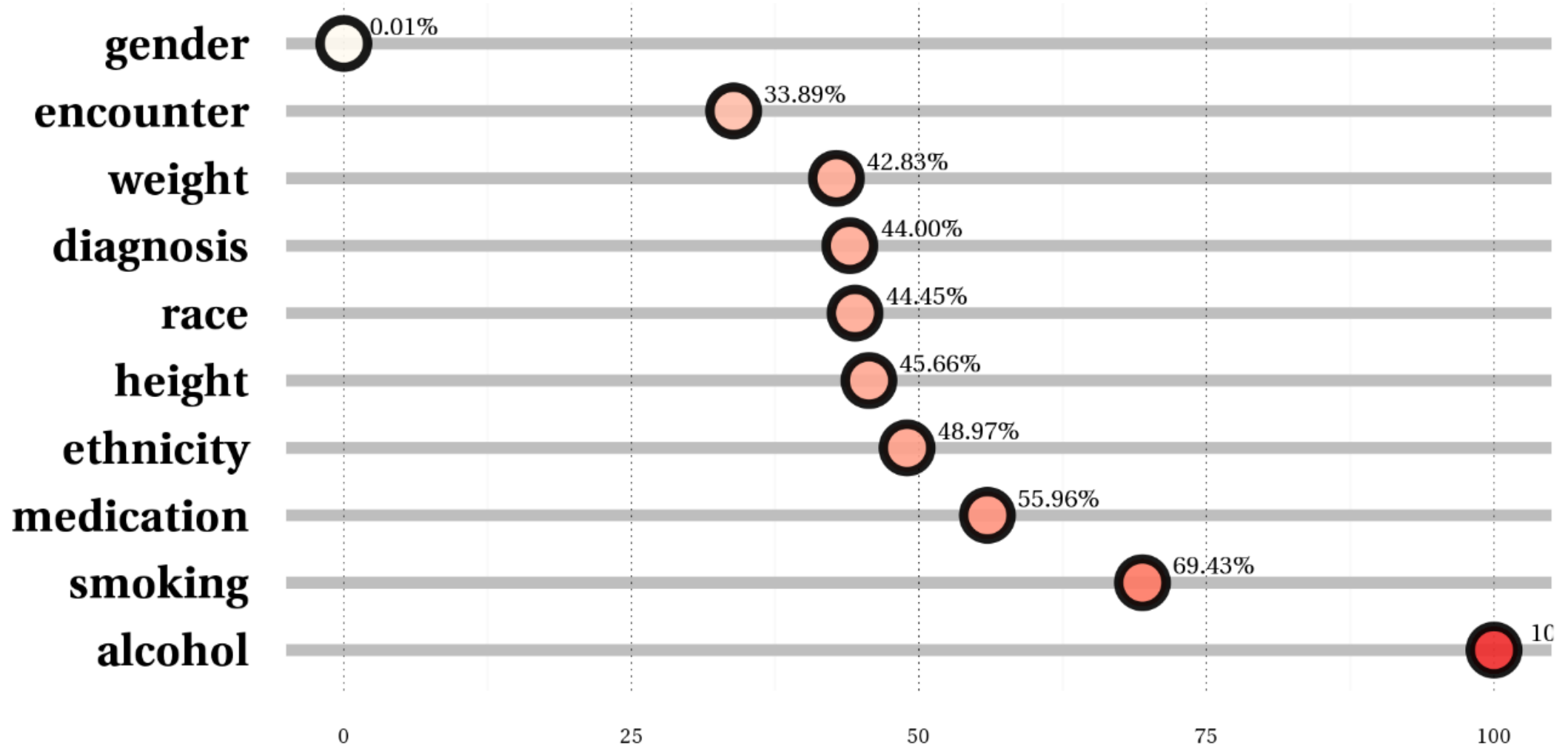


Fidelity example:

Detailing totals of key overlap across core tables

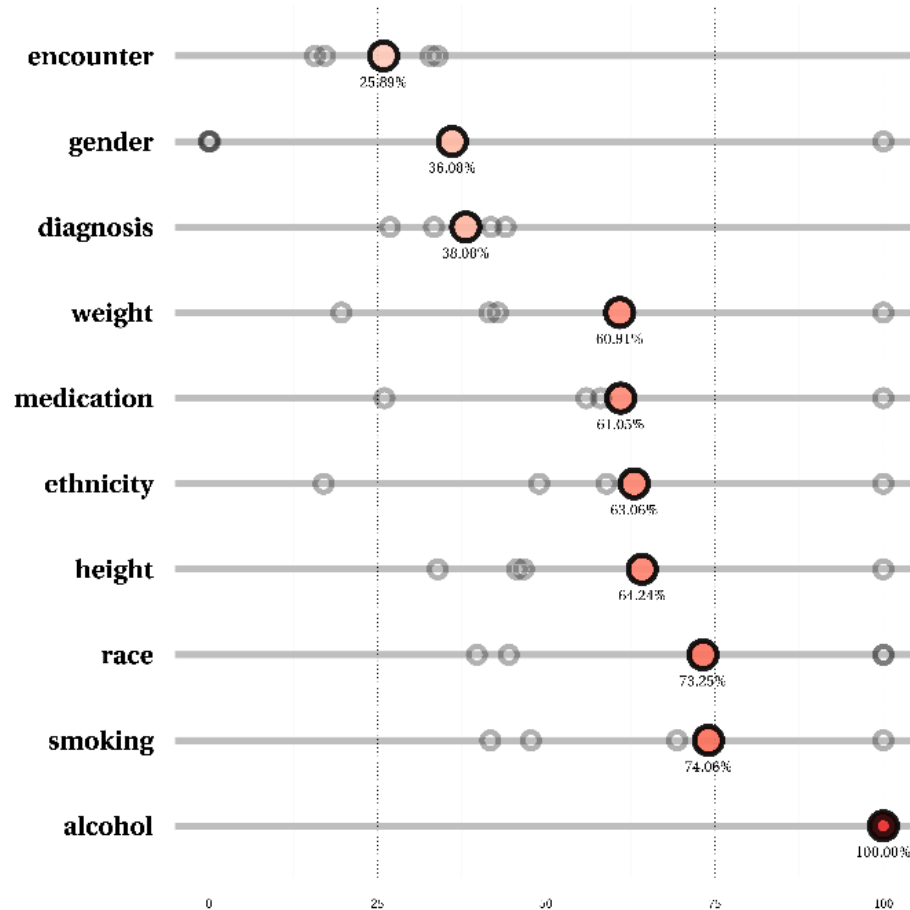


Completeness/Fidelity example: Percent of patients missing specific key clinical indicators

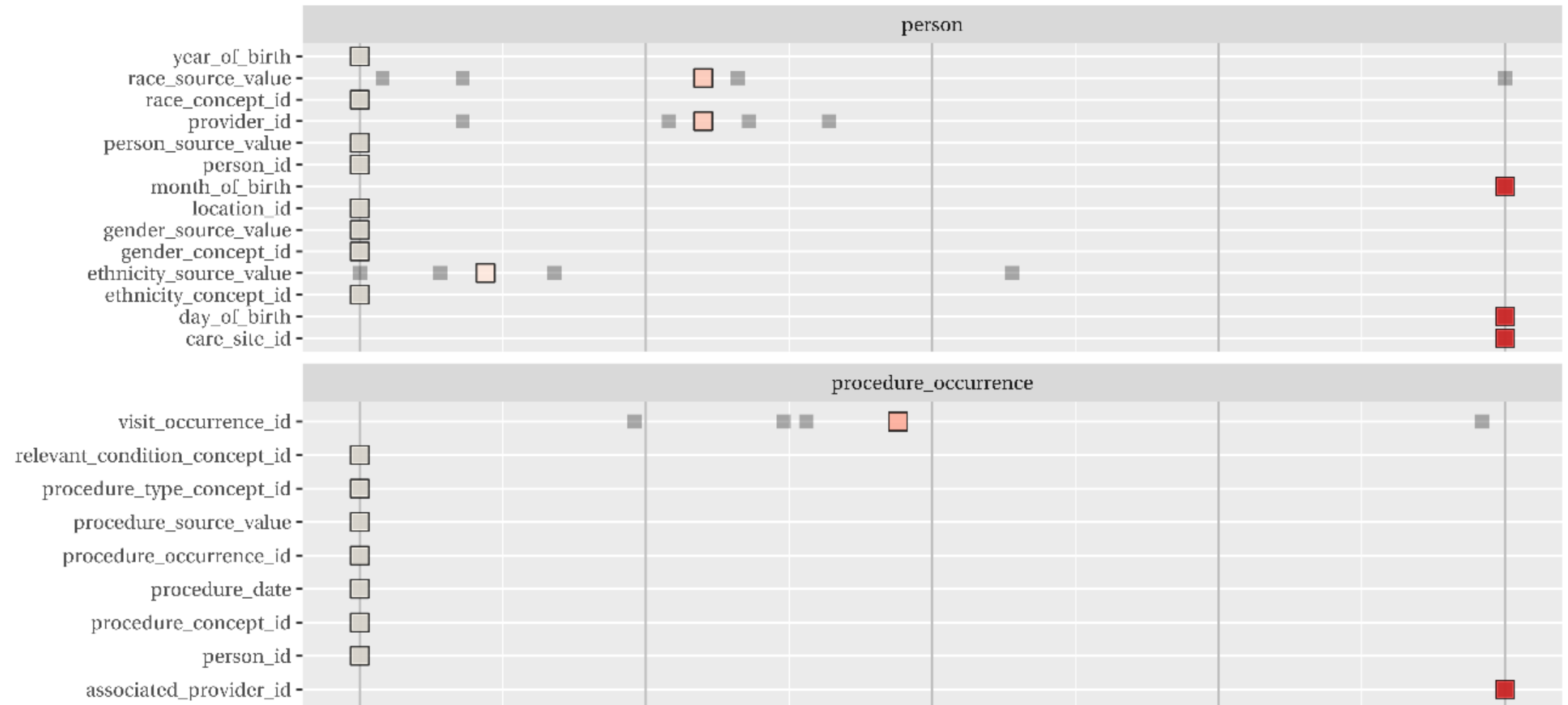


Completeness/Fidelity example across sites: Percent of patients missing specific key clinical indicators

Figure 2. Overall missingness in key indicators



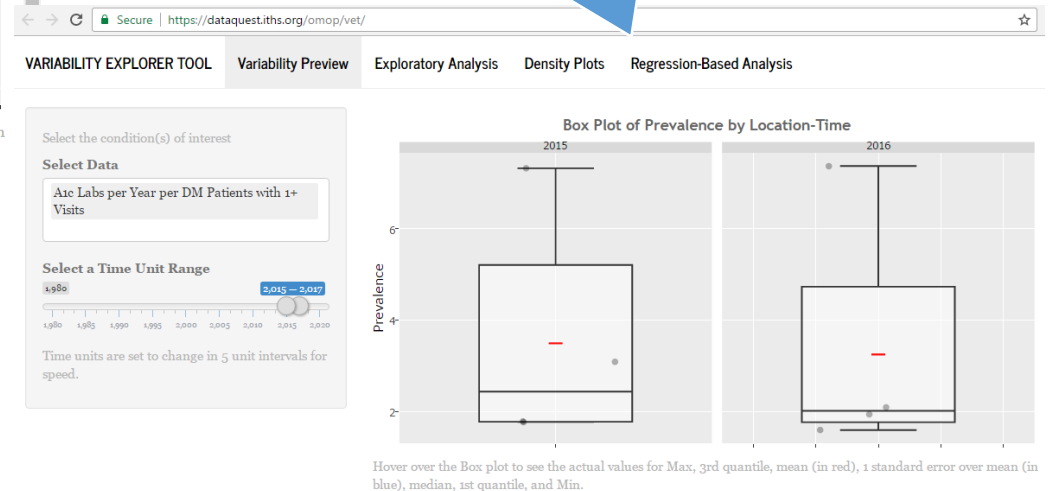
Completeness example across sites/clinics: Percent of patients missing in columns across sites



Plausability example across sites/clinics: # of Hemoglobin A1c's per year per diabetes patient with 1+ visit



Zoom to 2015-16



Next Steps

- Finalize SOP manual for DQe
- Iterate and refining functionality in DQe-v
- Create standard report of data quality findings
- Add new tests as needed...

Thank you!

Contact: Kari Stephens
kstephen@uw.edu

<https://dataquest.iths.org/>

<https://github.com/WWAMI-DataQuest>